Title
Plasmodium vivax From Duffy-Negative and Duffy-Positive Individuals Share Similar Gene Pools in East Africa

Journal
The Journal of Infectious Diseases, 224(8)

ISSN
0022-1899

Authors
Kepple, Daniel
Hubbard, Alfred
Ali, Musab M
et al.

Publication Date
2021-10-28

DOI
10.1093/infdis/jiab063

Peer reviewed

The Journal of Infectious Diseases

MAJOR ARTICLE

Plasmodium vivax From Duffy-Negative and Duffy-Positive Individuals Share Similar Gene Pools in East Africa

Daniel Kepple,[1] Alfred Hubbard,[2] Musab M. Ali,[3] Beka R. Abargero,[4] Karen Lopez,[2] Kareen Pestana,[1] Daniel A. Janies,[2] Guiyun Yan,[5] Muzamil Mahdi Hamid,[3] Delenasaw Yewhalaw,[4,6] and Eugenia Lo[1]

[1]Biological Sciences, University of North Carolina at Charlotte, Charlotte, North Carolina, USA, [2]Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, North Carolina, USA, [3]Department of Parasitology and Medical Entomology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan, [4]Tropical Infectious Disease Research Center, Jimma University, Jimma, Ethiopia, [5]Program in Public Health, University of California at Irvine, Irvine, California, USA, and [6]School of Medical Laboratory Sciences, Institute of Health, Jimma University, Jimma, Ethiopia

Plasmodium vivax malaria was thought to be rare in Africa, but an increasing number of *P. vivax* cases reported across Africa and in Duffy-negative individuals challenges this dogma. The genetic characteristics of *P. vivax* in Duffy-negative infections, the transmission of *P. vivax* in East Africa, and the impact of environments on transmission remain largely unknown. This study examined genetic and transmission features of *P. vivax* from 107 Duffy-negative and 305 Duffy-positive individuals in Ethiopia and Sudan. No clear genetic differentiation was found in *P. vivax* between the 2 Duffy groups, indicating between-host transmission. *P. vivax* from Ethiopia and Sudan showed similar genetic clusters, except samples from Khartoum, possibly due to distance and road density that inhibited parasite gene flow. This study is the first to show that *P. vivax* can transmit to and from Duffy-negative individuals and provides critical insights into the spread of *P. vivax* in sub-Saharan Africa.

Keywords. *Plasmodium vivax*; Duffy-negative; transmission pathways; East Africa; genetic diversity.

Plasmodium vivax malaria is a neglected tropical disease, despite being more geographically widespread than other forms of malaria [1], and causes 132–391 million clinical infections each year [2]. *P. vivax* was previously thought to be rare in Africa because people of African descent often lack the Duffy antigen-chemokine receptor (DARC). DARC is a glycoprotein on the surface of red blood cells that allows *P. vivax* to invade human erythrocytes [3]. However, recent studies have reported several cases of *P. vivax* in Duffy-negative people across Africa where Duffy-negative populations are predominant [4, 5]. It is as yet unclear whether these cases are merely infections or can develop into gametocytes and transmit from Duffy-negative to other individuals. The latter scenario constitutes a significant public health threat, especially considering antimalarial resistance [6] and the pathogen's unique ability to form dormant-stage hypnozoites in host liver, giving rise to relapses from months to years later and impacting malaria elimination progress [7, 8]. Compared to *Plasmodium falciparum*, *P. vivax* has a broader temperature tolerance and an earlier onset of gametocyte development [9], enabling it to spread through the diverse sub-Saharan African climate [10]. However, the effect of environments on *P. vivax* transmission has not been examined in African settings.

Ethiopia and Sudan are two African countries where *P. vivax* malaria is prevalent [11], and Duffy-negative and Duffy-positive individuals coexist [3]. *P. vivax* prevalence is lower in Duffy-negative than Duffy-positive individuals [4, 12]. In Central and West Africa, more than 97% of the population is Duffy-negative [3]. By contrast, in Ethiopia and Sudan, about 35% and 50% of the general population, as well as 20% and 18% of the hospitalized malaria patients, respectively, are Duffy negative [3, 9, 13]. Other East African countries such as Eritrea and Madagascar have also reported significant *P. vivax* infections in Duffy-positive and Duffy-negative individuals who coexist [3, 14], suggesting the potential for between-host transmission. The genetic diversity and population structure of *P. vivax* have been shown to be substantially different from *P. falciparum* [7, 15]. *P. vivax* has a higher nucleotide diversity than *P. falciparum* [15], likely due to frequent gene flow via human movement, higher transmission intensity, and recombination, and variation in host susceptibility [7]. In Papua New Guinea, *P. vivax* had a 3.5-fold higher rate of polyclonality and nearly double the multiplicity of infection (MOI) than *P. falciparum* infections [16]. Similarly, in Cambodia [17], the Indo-West Pacific [18, 19], and the Brazilian Amazon [20], *P. vivax* had a higher microsatellite diversity than its sympatric *P. falciparum*. These findings highlight the potential for *P. vivax* to adapt to a wide range of landscapes and climates.

The questions of whether *P. vivax* infecting Duffy-negative individuals is genetically similar to that from Duffy-positive

Received 18 September 2020; editorial decision 26 January 2021; accepted 29 January 2021; published online February 3, 2021.

Correspondence: Daniel Kepple, BS, Biological Sciences, University of North Carolina at Charlotte, 9201 University City Blvd, Woodward hall 367, Charlotte, NC 28223 (dkepple@uncc.edu).

The Journal of Infectious Diseases® 2021;224:1422–31

DOI: 10.1093/infdis/jiab063

individuals in the same region, and how landscape or environmental factors influence *P. vivax* gene flow in Africa, remain unclear. This study had 2 objectives: (1) to examine genetic variation and source-sink dynamics of *P. vivax* between Duffy-negative and Duffy-positive populations in Ethiopia and Sudan; and (2) to estimated how landscape and environmental factors influence parasite gene flow. Findings provide critical insights into whether and how *P. vivax* evolves and spreads in Duffy-negative individuals and highlight the need for *P. vivax* diagnosis, tracking, and control in Africa.

METHODS

Ethics Statement

Scientific and ethical clearance was obtained from the institutional scientific and ethical review boards of Jimma University, Ethiopia, the ethical committee of Institute of Endemic Diseases, Sudan (approval number 1/2014), the ethical committee of the State Ministry of Khartoum and Kassala State Ministry of Health, Sudan, and University of North Carolina, Charlotte, NC. Written informed consent/assent for study participation was obtained from all consenting heads of households, parents/guardians (for minors younger than 18 years), and each individual who was willing to participate in the study.

Study Areas and Genotyping

Clinical samples were obtained from 6 different study sites (3 from Ethiopia and 3 from Sudan). Study sites in Ethiopia were Jimma, Gojeb, and Arjo; and study sites in Sudan were Khartoum, River Nile, and New Halfa. These sites are located around the Ethiopia-Sudan border area (Figure 1) and experience low to moderate transmission [21]. Samples were screened for *P. vivax* by 18S-based quantitative polymerase chain reaction (PCR) (Supplementary Material) [12]. A total of 305 and 107 *P. vivax* samples from Duffy-positive and Duffy-negative individuals, respectively, were included in microsatellite analyses. These included 150 samples from Duffy-positive and 83 samples from Duffy-negative individuals in Ethiopia, and 155 and 24 samples, respectively, in Sudan (Table 1). Eight microsatellites, which mapped to 6 chromosomes, were typed for *P. vivax* following the published protocol (Supplementary Material) [9].

Linkage Disequilibrium and Multiplicity of Infections

To examine whether the microsatellite loci are in linkage disequilibrium, multilocus linkage disequilibrium was assessed among samples for each study site using LIAN version 3.7 [22]. The standardized index of association (I_A^S), which measures the strength of linkage disequilibrium and is viewed as a function of recombination rate among samples, was calculated. Due to the small number of samples from Duffy-negative individuals in New Halfa, Sudan (n = 2), this population was excluded from linkage disequilibrium analyses. The percentage of polyclonal infections and average MOI were estimated for each study site and Duffy group. MOI was scored as the maximum number of alleles observed in each sample, when all loci were considered.

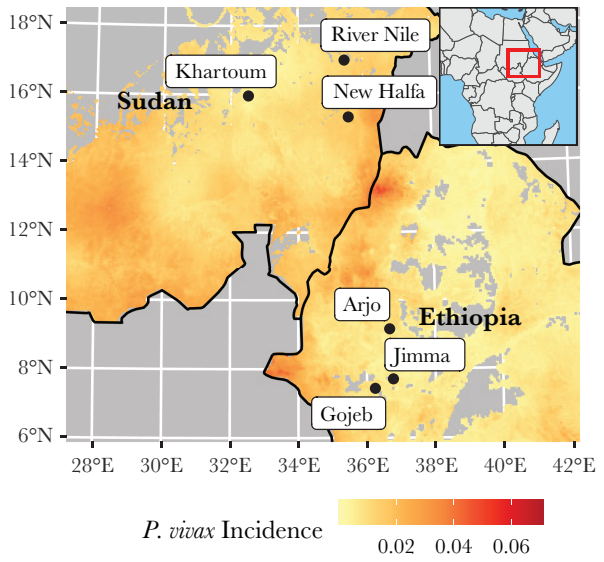

Figure 1. Study sites with *P. vivax* incidence heatmap for reference. Each point corresponds to 1 of the 6 sample locations, and the black lines represent the borders of Ethiopia and Sudan. The incidence layer was modeled by Battle et al [21] and represents the estimated cases per 1000 people in 2017.

Genetic Diversity and Population Structure

Genotypic variation was calculated in GENODIVE version 2.0b27 based on the mean of individual markers [23] and tested for significance by 2-tailed *t* tests. A model-based Bayesian method implemented in STRUCTURE version 2.3.4 [24] was performed separately at the site and country levels to examine the partitioning of individuals to genetic clusters (details in Supplementary Material). The partitioning of clusters was visualized with DISTRUCT [25]. In addition, discriminant analysis of principle components (DAPC) was performed to identify the optimal number of components that best separate samples into predefined clusters [26]. Phylogenetic analyses were conducted at the population level to examine the relatedness of *P. vivax* from Duffy-positive and Duffy-negative individuals (Supplementary Material).

Identification of Source-Sink Dynamics and Direction of Transmission

The StrainHub software [27] version 0.2.0 was used to generate a transmission network from the phylogeny of our samples. The squared Euclidean distances based on the number of times a certain allele is found in 2 individuals were computed between each pair of individuals using GENODIVE and an unrooted neighbor-joining tree was constructed using the ape R package version 5.4_1 [28]. StrainHub used the genetic relationships from the tree and the associated metadata states to reconstruct an estimate of ancestral states (Supplementary Material). Samples were grouped by country, site, and Duffy status to

Table 1. Host Duffy Status, Linkage Disequilibrium, and Complexity of Infection Among *Plasmodium vivax* Samples by Study Site

Site	Duffy Status	Sample Size	IA^S	No. of Polyclonal Infections (%)	MOI, Mean, Median (Range)
Ethiopia					
Arjo	Duffy positive	56	0.07*	4 (7.14)	1.08, 1 (1–2)
	Duffy negative	15	0.15*	2 (13.33)	1.13, 1 (1–2)
Gojeb	Duffy positive	62	0.04*	19 (30.65)	1.44, 1 (1–2)
	Duffy negative	4	0.26*	0	1, 1 (1)
Jimma	Duffy positive	32	0.07*	4 (12.50)	1.18, 1, (1–3)
	Duffy negative	64	0.03*	11 (17.19)	1.21, 1 (1–2)
Total	Duffy positive	150	0.03*	27 (18)	1.20, 1 (1–3)
	Duffy negative	83	0.04*	13 (15.66)	1.18, 1 (1–2)
Sudan					
Khartoum	Duffy positive	86	0.25*	5 (5.81)	1.06, 1 (1–2)
	Duffy negative	7	0.77*	0	1, 1 (1)
New Halfa	Duffy positive	36	0.04*	3 (8.33)	1.09, 1 (1–2)
	Duffy negative	2	NA	1 (50)	1.5, 1 (1–2)
River Nile	Duffy positive	33	0.08*	7 (21.21)	1.31, 1 (1–3)
	Duffy negative	15	0.12*	2 (13.33)	1.15, 1 (1–2)
Total	Duffy positive	155	0.12*	15 (9.68)	1.11, 1 (1–3)
	Duffy negative	24	0.20*	3 (12.50)	1.14, 1 (1–2)
All Sites	Duffy positive	305	0.05*	42 (13.77)	1.16, 1 (1–3)
	Duffy negative	107	0.04*	16 (14.95)	1.17, 1 (1–2)

Abbreviations: IA^S, standardized index of association; NA, not applicable; MOI, multiplicity of infection.

Asterisk denotes a significant level at $P < .05$.

illustrate the directions of transmission. Values near 1 indicate a source, values near 0 indicate a sink, and values in the middle suggest the node is a hub. It is worth noting that StrainHub is a phylogenetic method that does not account for recombination among lineages. To validate the results of StrainHub, relatedness among groups were also computed using the identical-by-descent method described in Taylor et al [29] (Supplementary Material).

Analyses of Geographic and Landscape Factors

To test for isolation by distance, the Mantel test was performed on the matrix of pairwise fixation index (F_{ST}) values. This test was conducted with the ade4 R package, version 1.7_15, with 10 000 permutations performed to estimate significance [30]. For this analysis, the geographical coordinates of the 6 study sites were used for estimating physical distances between sites, and *P. vivax* samples from Duffy-positive and Duffy-negative individuals were combined for each study site.

To determine whether the observed patterns of genetic differentiation are explained by landscape factors, resistance surfaces were fit to the genetic data for a variety of gridded, environmental datasets including: road density, as estimated by Meijer et al [31]; elevation, as measured by the Shuttle Radar Topography Mission [32]; and land cover, obtained from the MCD12Q1 dataset [33] (Supplementary Figure 3). Road density was selected as a proxy for human movement, whereas elevation and land cover were chosen to represent mosquito habitats. Parasite samples were gathered between 2017 and 2019, and so land cover data for 2018 was chosen as representative. The other 2 datasets are static and all 3 were projected to a consistent coordinate system (Supplementary Material). Resistance surfaces based on these 3 environmental datasets that best explain the observed genetic distances (pairwise F_{ST}-values) were modeled using the ResistanceGA R package version 4.0-14 [34] (Supplementary Material). Resistance surfaces were fit for each individual environmental variable, and composite resistance surfaces were fit for each possible combination of multiple environmental variables.

RESULTS

Linkage Disequilibrium and Multiplicity of Infections

Significant linkage disequilibrium was detected for all pairwise combinations of microsatellite loci (Bonferroni-corrected $P < .05$). When all loci were pooled together, *P. vivax* from Duffy-negative individuals in both countries showed a slightly higher level of linkage and/or rate of recombination than samples from Duffy-positive individuals (Table 1). I_A^S values of *P. vivax* from Duffy-negative individuals ranged from 0.03 (Jimma, Ethiopia) to 0.77 (Khartoum, Sudan), whereas I_A^S values ranged from 0.04 (Gojeb, Ethiopia) to 0.25 (Khartoum, Sudan) in Duffy-positive individuals.

P. vivax from Duffy-negative and Duffy-positive individuals showed a comparable rate of polyclonal infections (15% and 14%, respectively; Table 1). Among the samples from Duffy-negative individuals, the highest rate of polyclonal infections was observed in Jimma, Ethiopia (17.2%), and no polyclonal infections in Khartoum, Sudan and Gojeb, Ethiopia, although

there were few samples from Duffy-negative individuals in these locations. Among the samples from Duffy-positive individuals, the highest rate of polyclonal infections was observed in Gojeb, Ethiopia (30.7%), followed by River Nile, Sudan (21%); whereas the lowest was observed in Khartoum (5.8%; Table 1). MOI was also similar between the Duffy-negative (mean = 1.17) and Duffy-positive (mean = 1.16; Table 1) infections, with a maximum of 3 clones detected within a sample. Because we were unable to confidently differentiate the genotypes of the different clones in samples with more than 1 alleles in 2 or more loci, 16 of the polyclonal samples were discarded in the genetic analyses (Supplementary Table 1).

Genetic Diversity and Population Structure

In Ethiopia and Sudan, *P. vivax* in Duffy-positive individuals showed significantly higher genotypic and allelic diversity than in Duffy-negative individuals ($P < .05$; Table 2). Genetic diversity of *P. vivax* in Duffy-positive individuals was similar between Ethiopia and Sudan. However, the diversity of *P. vivax* in Duffy-negative individuals from Ethiopia was much higher than that from Sudan, likely due to the differences in sample size. Interestingly, genotypic evenness of *P. vivax* was higher in Duffy-negative than Duffy-positive infections in both countries, though it was not significant ($P = .076$; Table 2). Given that *P. vivax* in Duffy-negative and Duffy-positive individuals shared a similar set of alleles in each of the studied loci (Supplementary Table 2), we found no evidence of population bottleneck in both Duffy groups.

STRUCTURE analyses indicated 2 most probable genetic clusters among all samples at the site level (purple and yellow clusters; Figure 2). In Ethiopia, most of the samples from Arjo, Gojeb, and Jimma predominately belonged to the purple cluster. No distinguishable differences were detected between samples from Duffy-positive and Duffy-negative infections. In Sudan, however, a mixed clustering pattern was observed. About 80% of the samples from Duffy-positive individuals in Khartoum belonged to the yellow cluster (Figure 3), whereas those in New Halfa and River Nile had a mixture of yellow and purple clusters. When the samples were analyzed by country and Duffy status rather than by site, 3 clusters were indicated (purple, blue, and yellow; Supplementary Figure 1). In Ethiopia, *P. vivax* from Duffy-positive individuals had a majority of purple cluster and *P. vivax* from Duffy-negative individuals had both purple and blue clusters. In Sudan, *P. vivax* from Duffy-positive individuals had mostly the yellow cluster and *P. vivax* from Duffy-negative individuals had mixed yellow, blue, and purple clusters. No marked difference was observed between Duffy-positive and Duffy-negative individuals in respective countries. Similar to the STRUCTURE results, DAPC analyses showed that *P. vivax* from individuals in Khartoum, Sudan was distinct from the others (Figure 3). In general, *P. vivax* from Duffy-positive and Duffy-negative individuals in the same site were not clearly distinguishable, except for samples in Jimma and

Table 2. Comparison of Genetic Diversity Based on Microsatellite Markers in Duffy-Negative and Duffy-Positive *Plasmodium vivax* Infections

Study Site	Duffy Status	Sample Size	Genotypic Diversity			Gene Diversity	
			G	*D*	*E*	N_e	H_e
Ethiopia							
Arjo	Duffy positive	56	41	0.94	0.44	8.48	0.87
	Duffy negative	15	15	0.93	1	5.48	0.85
Bonga	Duffy positive	62	35	0.94	0.49	8.07	0.85
	Duffy negative	4	2	0.38	0.80	1.95	0.65
Jimma	Duffy positive	64	40	0.94	0.71	8.47	0.84
	Duffy negative	32	24	0.91	0.29	7.72	0.88
Total	Duffy positive	182	116	0.94	0.55	25	0.85
	Duffy negative	51	41	0.74	0.70	15	0.79
Sudan							
Khartoum	Duffy positive	86	39	0.86	0.27	5.65	0.81
	Duffy negative	7	3	0.57	1	2.20	0.63
New Halfa	Duffy positive	36	22	0.83	0.27	8.00	0.88
	Duffy negative	2	1	0	1	1.13	0.13
River Nile	Duffy positive	33	19	0.88	0.43	7.00	0.86
	Duffy negative	15	6	0.60	0.42	5.82	0.91
Total	Duffy positive	155	80	0.86	0.32	20.65	0.85
	Duffy negative	24	10	0.39	0.81	9.15	0.56
P value		…	…	.025*	.076	.009*	.105

Statistical tests were performed between Duffy-positive and Duffy-negative individuals of all study sites. Asterisks indicate *P* values significant at < .05.

Abbreviations: *D*, Simpson's diversity index corrected for sample size; *E*, genotypic evenness; *G*, number of multilocus genotypes corrected for sample size; *He*, expected heterozygosity corrected for sample size [13]; *Ne*, number of effective alleles [35].

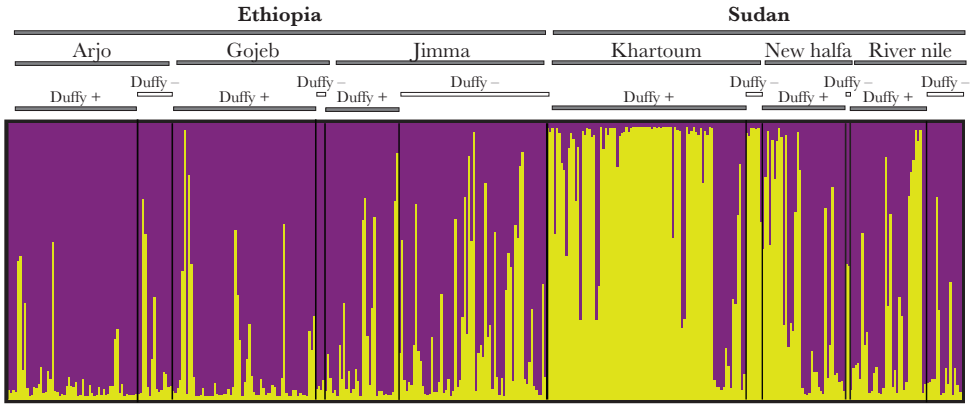

Figure 2. Bayesian inferences of the *K* cluster estimated by STRUCTURE for the *P. vivax* samples, according to host Duffy status and location. The most probable clusters are labeled by different colors and individuals are represented as columns.

Gojeb. In Jimma, samples from Duffy-negative individuals were slightly separated from Duffy-positive individuals. In Gojeb, the number of Duffy-negative samples was too small to indicate clear differentiations. Analyzes at the population levels also showed no clear distinction among samples by Duffy status or country (Supplementary Figure 2).

Source-Sink Dynamics and Direction of Transmission

The transmission networks produced by StrainHub showed that *P. vivax* from Duffy-positive and Duffy-negative individuals in Ethiopia served as the primary source, and *P. vivax* from the Sudanese Duffy-negative individuals was the sink population (Figure 4A). It should be noted that the weight of the arrows was

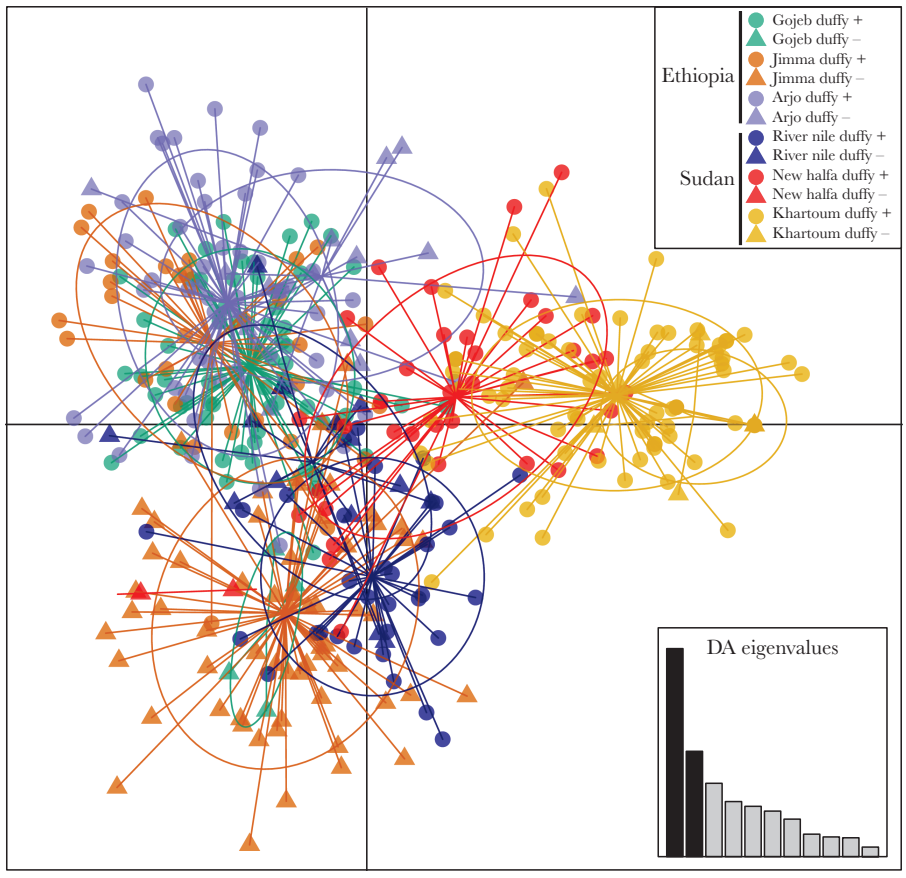

Figure 3. Results of discriminant analyses of principle components showing the first 2 principal components of the data grouped by site and host Duffy status. Dots are individuals and clusters are identified by symbol color, symbol shape, and inertia ellipses. Eigenvalues are shown in the inset.

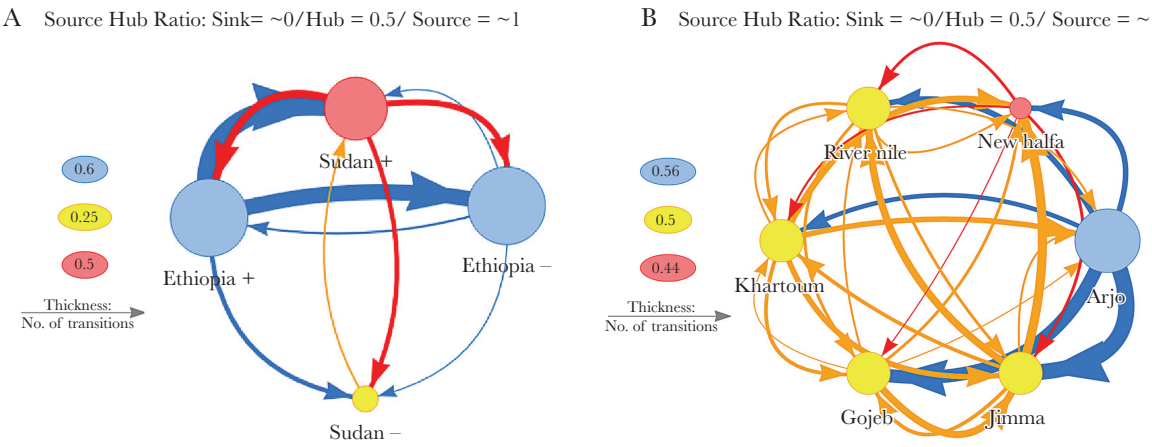

Figure 4. Source hub ratio (SHR) networks generated from (*A*) samples grouped by country and host Duffy status and (*B*) samples grouped by site. The size of each node is proportional to the node's SHR value. Node colors were randomly assigned to each unique SHR value present in the plot to improve readability. Similarly, arrow colors are paired with the color of the node from which they begin to aid with interpretation. The weight of each arrow is proportional to the number of transitions between its 2 nodes. The position of nodes is arbitrary and is not equivalent to the position of sample sites in geographic space.

not normalized by sample size. Although the Ethiopian Duffy-negative population had a high source hub ratio, the weight of the arrows leading from this population was not the thickest compared to Duffy-positive individuals, indicating that fewer transitions of *P. vivax* were from Duffy-negative individuals in Ethiopia. Based on identical-by-descent analyses, the average relatedness of *P. vivax* between Duffy-negative and Duffy-positive individuals was low but not significantly different to the values with the same Duffy status (Supplementary Table 3). Despite smaller samples from Duffy-negative individuals, these results suggested similar levels of transmission and recombination of *P. vivax* in individuals regardless of Duffy status and supported the StrainHub network showing that transition occurred both ways between Duffy-negative and Duffy-positive individuals (Figure 4A). At a site level, *P. vivax* from Arjo was a source, and *P. vivax* from infections in New Halfa was a sink (Figure 4B) for both phenotypes. The weight of the arrows suggest that Arjo had more transitions with the other populations in Ethiopia (Gojeb and Jimma) than those in Sudan.

Influence of Geographic and Landscape Factors

The Mantel test indicated a pattern of isolation by distance in the population F_{ST} values (r_M value = 0.325; *P* value = .017). The null hypothesis of no relationship between genetic and geographic distance was rejected. Based on the corrected Akaike information criteria (AICc) score and ranking, the resistance surface based on a single covariate that best fit the F_{ST} values was land cover (Figure 5A), while road density and elevation comprised the best fitting combination surface (Figure 5B). All resistance surfaces were significant (*P* value < .05; Supplementary Table 4), and the analysis showed that geographic distance alone does not explain the genetic distances in our dataset (*P* value > .025 for a 2-tailed test). ResistanceGA weighted elevation as 92% of the combination surface (Figure 5B), constituting most of the variation in resistance. Visual inspection suggests the exception was in the vicinity of Khartoum, where high resistance values were associated with high road density. On the other hand, land cover had the lowest AICc of any of the individual surfaces and all composite surfaces containing land cover were ranked close behind, with slightly higher AICc scores, suggesting land cover played an important role in explaining genetic distances, despite lower-ranked surface. Of the different land cover classes, grasslands and croplands were associated with low resistance, while savannas and open shrublands were modeled to have higher resistance. Taken together, the Mantel test and ResistanceGA analyses indicated that elevation and land cover were key factors in explaining *P. vivax* gene flow among our study sites in Ethiopia and Sudan, while geographic distance alone was less important.

DISCUSSION

Recent studies reported several cases of *P. vivax* infection in Duffy-negative people in different parts of Africa [4], including countries where Duffy-negative individuals are predominant [5]. The information on whether *P. vivax* is transmitted among Duffy-negative individuals and between Duffy-negative and Duffy-positive populations, and what landscape factors govern *P. vivax* transmission, is crucial to estimating the ability of *P. vivax* malaria to transmit and spread through Duffy-negative populations in Africa. This study provides critical insights into these questions using samples from Duffy-negative and Duffy-positive individuals in Ethiopia and Sudan. These countries were chosen because many *P. vivax* cases were recorded [13, 36] and Duffy-negative and Duffy-positive individuals coexist in the same areas [3, 11]. No clear differentiation was detected in *P. vivax* between Duffy-positive and Duffy-negative individuals for

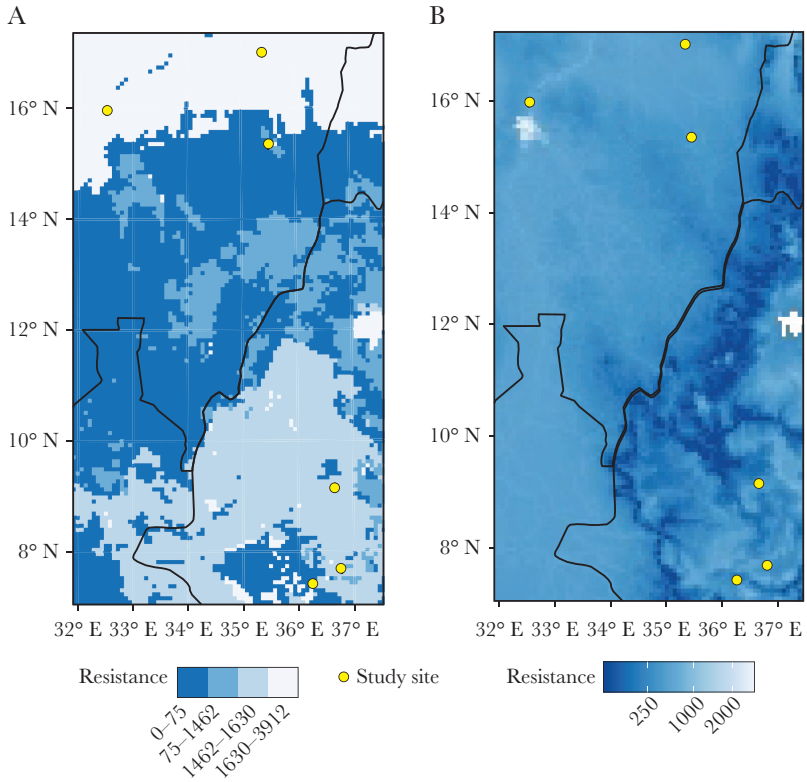

Figure 5. Resistance surfaces fit to genetic distances based on (*A*) land cover alone and (*B*) a combination of road density and elevation. Higher values indicate greater obstruction to gene flow. Study sites are denoted by points and the lines indicate the borders of Ethiopia and Sudan.

most of the studied sites. The StrainHub network indicated that transmission of *P. vivax* occurred from Duffy-positive to Duffy-negative individuals and vice versa. This suggests that Duffy-negative hosts are not a dead end for *P. vivax* that only result in infection and/or clinical symptoms, but that the parasite can also spread from one Duffy-negative host to another. *P. vivax* from Duffy-negative individuals in Sudan was identified as a sink population. While this might imply limited onward transmission, larger samples are needed to confirm this finding. Interestingly, our findings showed that the MOI was similar between the Duffy-negative and Duffy-positive infections. Based on microsatellites, about 9%–18% of the samples were polyclonal, consistent with our recent study that showed 18% (8 of 44) of *P. vivax* in Duffy-positive individuals from Ethiopia were polyclonal by whole-genome sequences [36]. Such moderate levels of polyclonality could be due to relapse clones [8] or submicroscopic infections, especially in Duffy-negative individuals with low parasitemia where different parasite clones may persist in the host for a long time without detection [37]. Additionally, transmission intensity, demographic factors, mosquito density, and vectoral capacity may contribute to the differences in MOI among study populations. Previous studies on *P. falciparum* have shown that areas with a higher abundance of mosquitoes, higher population density, and a younger age group are associated with a higher entomological inoculation rate [38]. These factors may apply to *P. vivax* in Duffy-negative and Duffy-positive populations, and consequently increase MOI.

The significantly lower genetic diversity observed in Duffy-negative compared to Duffy-positive infections suggested a lower transmission frequency and possibly and lower infectiousness amongst Duffy-negative Africans. It is noteworthy that the low number of Duffy-negative infections in Gojeb, Khartoum, and New Halfa might influence genetic diversity measures. While *P. vivax* from Duffy-negative individuals in Sudan was identified as a sink population and may imply limited onward transmission, further study is needed with larger samples to confirm these findings. To allow transmission in Duffy-negative individuals, *P. vivax* merozoites may have used non-Duffy receptors to invade human erythrocytes. Candidate genes for invasion, such as those coding for glycophosphatidylinositol-anchored micronemal antigen (GAMA) and tryptophan-rich antigens, have been shown to be expressed in patients with low parasite density. Also, CD71 (transferrin receptor 1, TfR1) has been shown to bind readily with the reticulocyte binding proteins based on in- vitro experiments [39]. These alternative ligand/receptor proteins may play a key role in erythrocyte invasion [5]. Further investigations at the whole-genome level are needed to uncover the Duffy-independent invasion pathway and transmission mechanism in Duffy-negative individuals [40].

Apart from the spread of *P. vivax* between Duffy-positive and Duffy-negative hosts, this study also reveals the potential of *P. vivax* to spread through the physical environments in East Africa. Parasite populations from Ethiopia and Sudan showed overlapping clustering patterns, except for Khartoum that was slightly different from the other samples. Low levels of linkage disequilibrium were observed among samples in most of our study sites, except for Khartoum which showed a high I_A^S-value, likely due to low parasite transmission in the capitol of Sudan, an urban setting with adequate local public health infrastructure [41]. While the Mantel test indicated that gene flow is associated with geographical distance, the resistance surface analysis suggested that landscape factors, including land cover and elevation, play a more important role in determining gene flow. The only exception was in Khartoum, where road density has been fit to a high resistance value, which likely explains the considerable genetic differentiation between Khartoum and other populations. The results suggest that *P. vivax* gene flow occurred more readily in the rural parts of Ethiopia and Sudan than in the more urban setting of Khartoum, consistent with other studies that showed higher diversity of *P. vivax* in areas close to the country border of Sudan and in agriculture areas outside the capitol [42]. These findings show that while sheer geographic distance serves as a barrier to gene flow, the specific features of the rural landscape in East Africa likely do not. Given that the Mantel test provides low analytical power [43] and there were few sample sites in this study, further investigations with additional sites, higher coverage of urban areas, and genome-wide data are needed to confirm these findings. Additionally, it might be informative to explore other environmental variables related to human activities and/or mosquito habitats, beyond the landscape data examined here.

In conclusion, this study is the first to show that *P. vivax* can transmit to and from Duffy-negative individuals and sheds light on the transmission pathways of Duffy-negative *P. vivax* in endemic regions of East Africa. We integrated genetic with landscape data and found that urban environments pose a greater hindrance to the transmission of *P. vivax* than rural and less-developed areas. The generally high genetic diversity of *P. vivax* in most study sites could be related to high transmission and frequent gene flow between Duffy-negative and Duffy-positive individuals as well as across the country border. Future investigations using whole genome sequence data with expanded *P. vivax* samples in East Africa are imperative to confirm these findings. Knowledge of *P. vivax* transmission mechanisms involving gametocyte infectivity experiments and in vitro invasion assays in Duffy-negative individuals would bring attention to the need for diagnosing and controlling *P. vivax* in Africa.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. We thank the field team from Jimma University and University of Khartoum for their technical assistance; the communities and hospitals for their support and willingness to participate in this research; Colton and Greyson Cantley, Haley Wheeler, Miriam Rodriguez, and Rahm Tawkaliyar for assistance with DNA extraction; and Dr Colby Ford for providing guidance on StrainHub.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institutes of Health.

Financial support. This work was supported by the National Institutes of Health (grant numbers R15 AI138002, R01 AI050243, U19 AI129326, and D43 TW001505).

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization (WHO). World malaria report 2018, **2018**. Geneva, Switzerland: WHO.
2. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. *Vivax* malaria: neglected and not benign. Am J Trop Med Hyg **2007**; 77:79–87.
3. Howes RE, Patil AP, Piel FB, et al. The global distribution of the Duffy blood group. Nat Commun **2011**; 2:266.
4. Albsheer MMA, Pestana K, Ahmed S, et al. Distribution of Duffy phenotypes among *Plasmodium vivax* infections in Sudan. Genes **2019**; 10:437.
5. Gunalan K, Niangaly A, Thera MA, Doumbo OK, Miller LH. *Plasmodium vivax* infections of Duffy-negative erythrocytes: historically undetected or a recent adaptation? Trends Parasitol **2018**; 34:420–9.
6. Anantabotla VM, Antony HA, Parija SC, et al. Polymorphisms in genes associated with drug resistance of *Plasmodium vivax* in India. Parasitol Int **2019**; 70:92–7.
7. Auburn S, Benavente ED, Miotto O, et al. Genomic analysis of a pre-elimination Malaysian *Plasmodium vivax* population reveals selective pressures and changing transmission dynamics. Nat Commun **2018**; 9:2585.
8. Markus MB. The hypnozoite concept, with particular reference to malaria. Parasitol Res **2011**; 108:247–52.
9. Lo E, Hemming-Schroeder E, Yewhalaw D, et al. Transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum* in Ethiopia and prevalence of antimalarial resistant genotypes. PLoS Negl Trop Dis **2017**; 11:e0005806.

10. Elgoraish AG, Elzaki SEG, Ahmed RT, et al. Epidemiology and distribution of *Plasmodium vivax* malaria in Sudan. Trans R Soc Trop Med Hyg **2019**; 113:517–24.
11. Sibley CH. A solid beginning to understanding *Plasmodium vivax* in Africa. J Infect Dis **2019**; 220:1716–8.
12. Lo E, Yewhalaw D, Zhong D, et al. Molecular epidemiology of *Plasmodium vivax* and *Plasmodium falciparum* malaria among Duffy-positive and Duffy-negative populations in Ethiopia. Malar J **2015**; 14:84.
13. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **1978**; 89:583–90.
14. World Health Organization (WHO). World Malaria Report 2015, **2015**. Geneva, Switzerland: WHO.
15. Hupalo DN, Luo Z, Melnikov A, et al. Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. Nat Genet **2016**; 48:953–8.
16. Fola AA, Harrison GLA, Hazairin MH, et al. Higher complexity of infection and genetic diversity of *Plasmodium vivax* than *Plasmodium falciparum* across all malaria transmission zones of Papua New Guinea. Am J Trop Med Hyg **2017**; 96:630–41.
17. Orjuela-Sánchez P, Sá JM, Brandi MC, et al. Higher microsatellite diversity in *Plasmodium vivax* than in sympatric *Plasmodium falciparum* populations in Pursat, Western Cambodia. Exp Parasitol **2013**; 134:318–26.
18. Gray KA, Dowd S, Bain L, et al. Population genetics of *Plasmodium falciparum* and *Plasmodium vivax* and asymptomatic malaria in Temotu Province, Solomon Islands. Malar J **2013**; 12:429.
19. Jennison C, Arnott A, Tessier N, et al. *Plasmodium vivax* populations are more genetically diverse and less structured than sympatric *Plasmodium falciparum* populations. PLoS Negl Trop Dis **2015**; 9:e0003634.
20. Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, Hartl DL. Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia. J Infect Dis **2007**; 195:1218–26.
21. Battle KE, Lucas TCD, Nguyen M, et al. Mapping the global endemicity and clinical burden of *Plasmodium vivax*, 2000–17: a spatial and temporal modelling study. Lancet **2019**; 394:332–43.
22. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics **2000**; 16:847–9.
23. Meirmans PG, Tienderen PHV. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes **2004**; 4:792–4.
24. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics **2000**; 155:945–59.
25. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes **2003**; 4:137–8.
26. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet **2010**; 11:94.
27. de Bernardi Schneider A, Ford CT, Hostager R, et al. StrainHub: a phylogenetic tool to construct pathogen transmission networks. Bioinformatics **2020**; 36:945–7.
28. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics **2019**; 35:526–8.
29. Taylor AR, Jacob PE, Neafsey DE, Buckee CO. Estimating relatedness between malaria parasites. Genetics **2019**; 212:1337–51.
30. Dray S, Dufour A, Chessel D. The ade4 package—II: two-table and K-table methods. R News **2007**; 7:47–52.
31. Meijer JR, Huijbregts MAJ, Schotten KCGJ, Schipper AM. Global patterns of current and future road infrastructure. Environ Res Lett **2018**; 13:064006.
32. Kobrick M, Crippen R: National Aeronautics and Space Administration (NASA) Earth Observing System Data and Information System Land Processes Distributed Active Archive Center. NASA shuttle radar topography mission global 1 arc second, **2013**. https://doi.org/10.5067/MEaSUREs/SRTM/SRTMGL1.003. Accessed 4 February 2021.
33. Friedl M, Sulla-Menashe D: National Aeronautics and Space Administration Earth Observing System Data and Information System Land Processes Distributed Active Archive Center. MCD12Q1 MODIS/Terra+Aqua land cover type yearly L3 Global 500 m SIN Grid V006, **2019**. https://doi.org/10.5067/MODIS/MCD12Q1.006. Accessed 4 February 2021.
34. Peterman WE. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. Methods Ecol Evol **2018**; 9:1638–47.
35. Nielsen R, Tarpy DR, Reeve HK. Estimating effective paternity number in social insects and the effective number of alleles in a population. Molecular Ecology **2003**; 12:3157–64.
36. Ford A, Kepple D, Abagero BR, et al. Whole genome sequencing of *Plasmodium vivax* isolates reveals frequent sequence and structural polymorphisms in erythrocyte binding genes. PLoS Negl Trop Dis **2020**; 14:e0008234.
37. Kasehagen LJ, Mueller I, Kiniboro B, et al. Reduced *Plasmodium vivax* erythrocyte infection in PNG Duffy-negative heterozygotes. PLoS One **2007**; 2:e336.
38. Vafa M, Troye-Blomberg M, Anchang J, Garcia A, Migot-Nabias F. Multiplicity of *Plasmodium falciparum* infection

in asymptomatic children in Senegal: relation to transmission, age and erythrocyte variants. Malar J **2008**; 7:17.
39. Chan LJ, Dietrich MH, Nguitragool W, Tham WH. *Plasmodium vivax* reticulocyte binding proteins for invasion into reticulocytes. Cell Microbiol **2020**; 22:e13110.
40. Cheng Y, Ahammed GJ, Yu J, et al. Putative WRKYs associated with regulation of fruit ripening revealed by detailed expression analysis of the *WRKY* gene family in pepper. Sci Rep **2016**; 6:39000.
41. Robert V, Macintyre K, Keating J, et al. Malaria transmission in urban sub-Saharan Africa. Am J Trop Med Hyg **2003**; 68:169–76.
42. Talha AA, Pirahmadi S, Mehrizi AA, Djadid ND, Nour BY, Zakeri S. Molecular genetic analysis of *Plasmodium vivax* isolates from Eastern and Central Sudan using *pvcsp* and *pvmsp-3α* genes as molecular markers. Infect Genet Evol **2015**; 32:12–22.
43. Legendre P, Fortin MJ, Borcard D. Should the Mantel test be used in spatial analysis? Methods Ecol Evol **2015**; 6:1239–47.